

B.D.
6/7/02

**Prioritätsbescheinigung über die Einreichung
einer Patentanmeldung**

Aktenzeichen:

101 06 580.9

Anmeldetag:

13. Februar 2001

Anmelder/Inhaber:

Philips Corporate Intellectual Property GmbH,
Hamburg/DE

Bezeichnung:

Spracherkennungssystem, Trainingseinrichtung
und Verfahren zum Berechnen von Iterationswerten
für freie Parameter eines Maximum-Entropie-
Sprachmodells

IPC:

G 10 L 15/06

**CERTIFIED COPY OF
PRIORITY DOCUMENT**

Die angehefteten Stücke sind eine richtige und genaue Wiedergabe der
ursprünglichen Unterlagen dieser Patentanmeldung.

München, den 2. November 2001
Deutsches Patent- und Markenamt
Der Präsident
Im Auftrag

Waasmaier

ZUSAMMENFASSUNG

Spracherkennungssystem, Trainingseinrichtung und Verfahren zum Berechnen von Iterationswerten für freie Parameter eines Maximum-Entropie-Sprachmodells

Die Erfindung betrifft ein Spracherkennungssystem und ein Verfahren zum Berechnen von Iterationswerten für freie Parameter $\lambda\alpha$ des Maximum-Entropie-Sprachmodells. Es ist im Stand der Technik bekannt, diese freien Parameter $\lambda\alpha$ z.B. mit Hilfe eines GIS-Trainingsalgorithmus zyklisch iterativ zu approximieren. Zyklisch bedeutet in diesem Fall, dass bei jedem Iterationsschritt n eine zyklisch vorbestimmte Merkmalsgruppe $A_i(n)$ des Sprachmodells zur Berechnung des $n+1$ Iterationswertes für die freien Parameter ausgewertet wird. Eine derartig starr zyklisch zugeordnete Merkmalsgruppe $A_i(n)$ ist jedoch nicht immer am besten geeignet, den GIS-Trainingsalgorithmus in einer aktuellen Situation am schnellsten und effektivsten konvergieren zu lassen. Es wird deshalb erfindungsgemäß ein Verfahren zur Auswahl der in dieser Hinsicht am besten geeigneten Merkmalsgruppe vorgeschlagen, wobei der Grad der Anpassung von Iterationsrandwerten $m_\alpha^{(n)}$ an jeweils zugehörige gewünschte Randwerte $m\alpha$ für alle Merkmale der jeweiligen Merkmalsgruppe als Kriterium für die Auswahl der Merkmalsgruppe dient.

Fig. 1

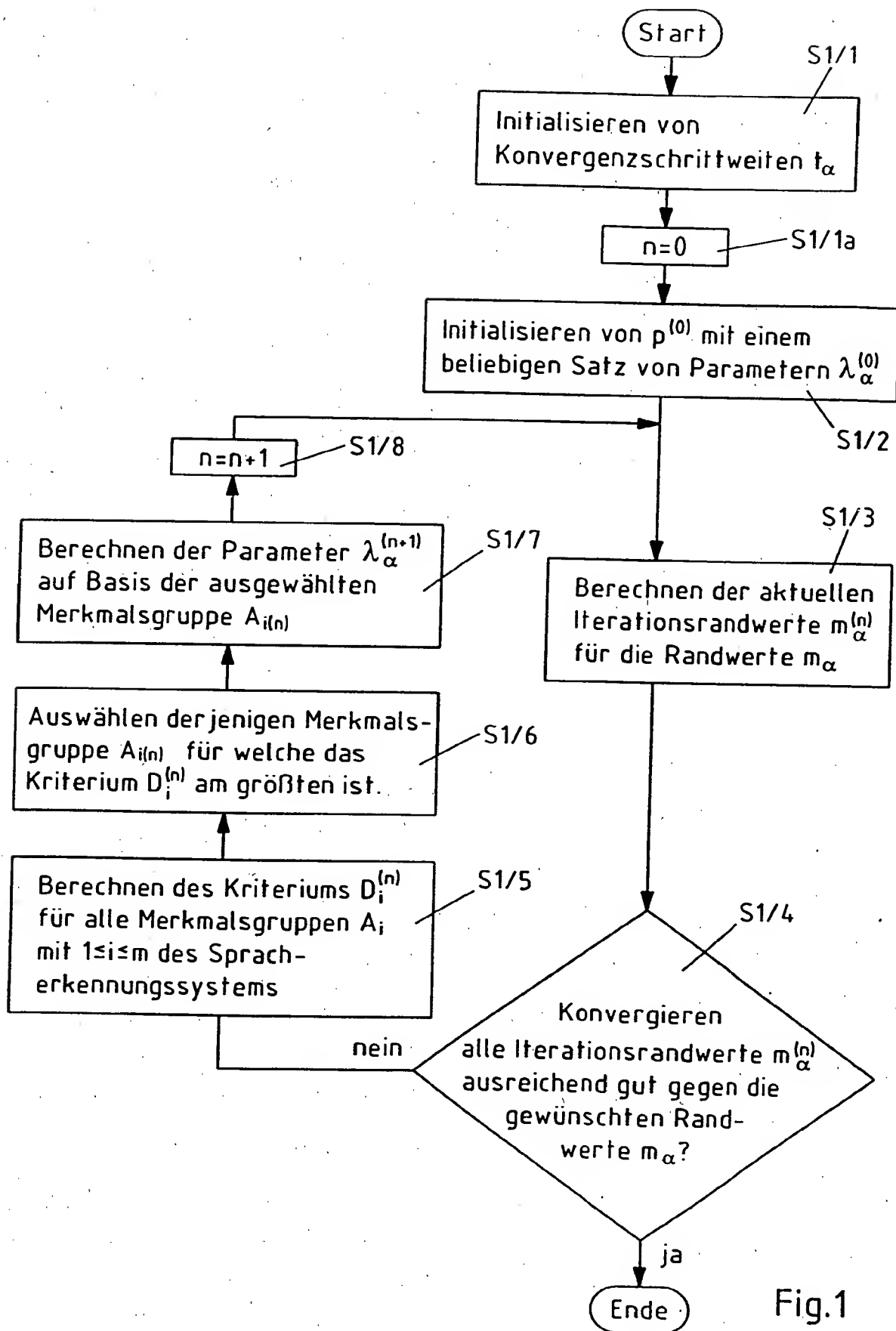


Fig.1

BESCHREIBUNG

Spracherkennungssystem, Trainingseinrichtung und Verfahren zum Berechnen von Iterationswerten für freie Parameter eines Maximum-Entropie-Sprachmodells

Die Erfindung betrifft ein Verfahren zum Berechnen von Iterationswerten für freie Parameter λ des Maximum-Entropie (ME)-Sprachmodells gemäß dem Oberbegriff des Patentanspruchs 1.

Die Erfindung betrifft weiterhin ein Spracherkennungssystem und eine Trainingseinrichtung, in welchen ein derartiges Verfahren realisiert ist.

10

Im Stand der Technik ist es bekannt, dass in einem (ME)-Sprachmodell sog. freie Parameter λ definiert bzw. trainiert werden müssen. Ein bekannter Algorithmus zum Trainieren dieser freien Parameter λ ist der sog. Generalized Iterative Scaling (GIS)-Trainingsalgorithmus. Von diesem GIS-Trainingsalgorithmus sind mehrere Varianten bekannt; die vorliegende Erfindung betrifft allerdings nur eine sog. zyklische Variante, wonach freie Parameter λ iterativ wie folgt berechnet werden:

$$\lambda_{\alpha}^{(n+1)} = \lambda_{\alpha}^{(n)} + t_{\alpha} \cdot \log \left(\frac{m_{\alpha}}{m_{\alpha}^{(n)}} \cdot \frac{1 - \sum_{\beta \in A_{n(\bmod m)}} t_{\beta} \cdot m_{\beta}^{(n)}}{1 - \sum_{\beta \in A_{n(\bmod m)}} t_{\beta} \cdot m_{\beta}} \right) \quad (1)$$

20 Bei dieser zyklischen Variante ist jedem Iterationswert n eine Merkmalsgruppe A_i mit $i = n(\bmod m)$ von insgesamt m Merkmalsgruppen in dem Sprachmodell zugeordnet und die Iterationswerte $\lambda_{\alpha}^{(n+1)}$ werden jeweils für alle Merkmale α aus der aktuell zugeordneten Merkmalsgruppe A_i berechnet, bevor der Iterationsparameter n um 1 erhöht wird. Diese zyklische Variante des GIS-Trainingsalgorithmus ist z.B. offenbart in J.N. Darroch and D. Ratcliff "Generalized iterative scaling for log linear models", Annals Math. Stat., 25 43(5):1470-1480, 1972.

In der Formel (1) bezeichnet:

- n : den Iterationsparameter;
- m : die Anzahl der insgesamt in dem Sprachmodell vordefinierten
- 5 Merkmalsgruppen;
- $A_n(\text{mod } m)$: die dem Iterationsparameter n aktuell zugeordnete
- Merkmalsgruppe;
- α : ein bestimmtes Merkmal aus der Merkmalsgruppe $A_n(\text{mod } m)$;
- β : alle Merkmale aus der Merkmalsgruppe
- 10 $A_n(\text{mod } m)$;
- $\lambda_\alpha^{(n)}$: den n 'ten Iterationswert für den freien Parameter λ_α ;
- $\tau\alpha, \tau\beta$: Konvergenzschrittweiten;
- $m\alpha, m\beta$: gewünschte Randwerte in dem Sprachmodell; und
- $m_\alpha^{(n)}, m_\beta^{(n)}$: n 'te Iterationsrandwerte für die
- 15 gewünschten Randwerte $m\alpha$ bzw. $m\beta$.

Einige der soeben vorgestellten Parameter in der Formel (1) werden nachfolgend näher erläutert:

- 20 Der zyklischen Variante des GIS-Trainingsalgorithmus gemäß Formel (1) liegt der Gedanke zugrunde, dass alle in dem ME-Sprachmodell vorbestimmten Merkmale einzelnen Merkmalsgruppen A_i zugeordnet sind, von denen insgesamt m in dem Sprachmodell definiert sind. Ein Beispiel für ein Sprachmodell mit insgesamt $m = 3$ vordefinierten Merkmalsgruppen A_i mit $i = 1 \dots 3$ ist in Fig. 5 anschaulich dargestellt. Merkmale
- 25 können dabei allgemein einzelne Worte, Wortfolgen, Wortklassen, Folgen von Wortklassen oder komplexere Muster bezeichnen. So umfasst in Fig. 5 die Merkmalsgruppe A_1 Worte, z.B. das Wort "House" und Wortfolgen, z.B. "The Green". Demgegenüber umfasst die Merkmalsgruppe A_3 einzelne Wortklassen, z.B. "Adjektive" oder "Substantive" und Folgen von Wortklassen, z.B. "Adverb - Verb".

Bei der bekannten zyklischen Berechnung der freien Parameter $\lambda\alpha$ gemäß Formel (1) ist jedem Iterationsparameter n eine Merkmalsgruppe $A_i = A_n(\text{mod } m)$, d.h. gemäß modulo m fest zugeordnet. Diese starre zyklische Zuordnung hat folgenden Nachteil:

- 5 Sie lässt keinen Raum für eine gezielte Anpassung des GIS-Trainingsalgorithmus auf diejenigen Merkmalsgruppen, bei welchen noch ein großer Korrekturbedarf besteht. So kann es passieren, dass bei einem nachfolgenden Iterationsschritt iterative Randwerte, welche bei einem vorherigen Iterationsschritt bereits gut an den zugeordneten gewünschten Randwert angepasst waren, keine großen λ -Korrekturen benötigen. Die Korrektur anderer Parameter
10 wäre hier von Vorteil.

Es werden deshalb bei der traditionellen zyklischen Variante unnötig viele Iterationsschritte durchgeführt, um eine gute Abschätzung der gewünschten Randwerte und der gewünschten freien Parameter λ zu erzielen.

- 15 Ausgehend von diesem Stand der Technik ist es die Aufgabe der Erfindung, ein Spracherkennungssystem, eine Trainingseinrichtung und ein Verfahren zum Berechnen von Iterationswerten für freie Parameter λ des ME-Sprachmodells derart weiterzubilden, dass die iterative Berechnung effektiver und schneller wird.

- 20 Diese Aufgabe wird durch das in Patentanspruch 1 beanspruchte Verfahren gelöst.

- Demnach wird die Aufgabe dadurch gelöst, dass dem aktuellen Iterationsparameter n jeweils diejenige Merkmalsgruppe $A_i(n)$ mit $1 \leq i(n) \leq m$ zugeordnet wird, für welche,
25 gemäß einem vorbestimmten Kriterium, die Anpassung der Iterationswerte $m^{(n)}\alpha$ an die jeweils zugehörigen gewünschten Randwerte $m\alpha$ am schlechtesten im Vergleich zu allen m Merkmalsgruppen des Sprachmodells erfüllt ist.

- Aufgrund dieser erfindungsgemäßen Zuordnung von Merkmalsgruppen $A_i(n)$ zu einzelnen
30 Iterationsparametern/ Iterationsschritten n wird ein besseres Konvergenzverhalten des GIS-Trainingsalgorithmus zur Approximation der freien Parameter λ erzielt. Die iterative

- Berechnung der freien Parameter λ ist jetzt nicht mehr als zyklisch zu bezeichnen, weil die Zuordnung von Merkmalsgruppen $A_i(n)$ zu dem Iterationsparameter n nicht mehr zyklisch, sondern gemäß einem individuell zu berechnenden Kriterium erfolgt. Im Vergleich zur zyklischen Version hat diese erfindungsgemäße azyklische Berechnung den Vorteil einer effektiveren und schnelleren Berechnung der gewünschten Iterationswerte für die freien Parameter λ .

- Gemäß dem in Patentanspruch 2 beschriebenen ersten Ausführungsbeispiel der Erfindung erfolgt vor jeder Erhöhung des Iterationsparameters n die Berechnung des Kriteriums zur Auswahl der am besten für den Iterationsparameter n geeigneten Merkmalsgruppe A_i gemäß folgender Gleichung:

$$D_i^{(n)} = \left[\sum_{\alpha \in A_i} t_\alpha \cdot m_\alpha \log \left(\frac{m_\alpha}{m_\alpha^{(n)}} \right) + \left(1 - \sum_{\alpha \in A_i} t_\alpha \cdot m_\alpha \right) \log \left(\frac{1 - \sum_{\alpha \in A_i} t_\alpha \cdot m_\alpha}{1 - \sum_{\alpha \in A_i} t_\alpha \cdot m_\alpha^{(n)}} \right) \right]$$

- Der Index der ausgewählten Merkmalsgruppe wird anschließend folgendermaßen festgelegt:

$$i(n) = \arg \max_j D_j^{(n)}$$

- Vorteilhafterweise hat der GIS-Trainingsalgorithmus zur iterativen Berechnung der freien Parameter λ - und damit die mathematische Funktion $G()$ in Patentanspruch 1 -folgende Gestalt:

$$\lambda_\alpha^{(n+1)} = G() = \lambda_\alpha^{(n)} + t_\alpha \cdot \log \left(\frac{m_\alpha}{m_\alpha^{(n)}} \cdot \frac{1 - \sum_{\beta \in A_i(n)} t_\beta \cdot m_\beta^{(n)}}{1 - \sum_{\beta \in A_i(n)} t_\beta \cdot m_\beta} \right), \quad (1a)$$

- wobei dieser Algorithmus im wesentlichen aus dem Stand der Technik bekannt ist und oben als Formel (1) beschrieben wurde. Wie in der zyklischen Version werden die freien Parameter λ_α gemäß Formel 1a angepasst. Hierbei werden alle Merkmale α der ausgewählten Gruppe $A_i(n)$ behandelt.

Es ist vorteilhaft, wenn sowohl zur Berechnung des Kriteriums $D_i^{(n)}$ wie auch zur Berechnung der freien Parameter λ gemäß dem GIS-Trainingsalgorithmus eine spezielle Merkmalsfunktion f_α , nämlich vorzugsweise eine orthogonalisierte Merkmalsfunktion f_α^{ortho} verwendet wird.

5

Die Verwendung der orthogonalisierten Merkmalsfunktion f_α^{ortho} bewirkt generell eine Verbesserung der Konvergenzgeschwindigkeit des GIS-Trainingsalgorithmus. Durch Verwendung der orthogonalisierten Merkmalsfunktion bei dem erfindungsgemäßen Verfahren ergibt sich eine zusätzlich erhöhte Konvergenzgeschwindigkeit für den GIS-

10 Trainingsalgorithmus.

Weitere vorteilhafte Ausgestaltungen und Verwendungen des erfindungsgemäßen Verfahrens sind Gegenstand der Unteransprüche.

15 Die Aufgabe der Erfindung wird weiterhin durch ein Spracherkennungssystem und durch eine Trainingseinrichtung auf Basis des Maximum-Entropie-Sprachmodells gemäß den Patentansprüchen 8 und 9 gelöst. Die Vorteile dieses Spracherkennungssystems und der Trainingseinrichtung entsprechen den Vorteilen, wie sie oben für das erfindungsgemäße Verfahren diskutiert wurden.

20

Es erfolgt eine detaillierte Beschreibung eines Ausführungsbeispiels der Erfindung unter Bezugnahme auf die folgenden Figuren, wobei

Fig. 1 ein Flussdiagramm zur Berechnung des Kriteriums zur Auswahl einer geeigneten Merkmalsgruppe $A_i(n)$ für einen Iterationsparameter n gemäß der vorliegenden Erfindung;

25

Fig. 2 ein Verfahren zur Berechnung eines verbesserten orthogonalisierten Randwertes

$$m_\alpha^{ortho};$$

30

Fig. 3 ein Spracherkennungssystem gemäß der vorliegenden Erfindung; und

Fig. 4 ein Beispiel für Merkmalsgruppen in einem Sprachmodell (Stand der Technik)

5 beschreibt.

Fig. 1 veranschaulicht die einzelnen Verfahrensschritte eines erfindungsgemäßen Verfahrens zur Auswahl derjenigen Merkmalsgruppe $A_i(n)$, welche zur Berechnung von Iterationswerten $\lambda_\alpha^{(n+1)}$ gemäß dem GIS-Trainingsalgorithmus am besten geeignet ist.

10

Das Verfahren gemäß Fig. 1 sieht vor, dass in einem ersten Verfahrensschritt S1/1 zunächst Konvergenzschriftweiten α initialisiert werden müssen. In Schritt S1/1a wird der Iterationsparameter $n = 0$ gesetzt.

15 Weiterhin ist die Wahrscheinlichkeit $p(0)$ mit einem beliebigen Satz von Anfangsparametern $\lambda_\alpha^{(0)}$ zu initialisieren. Dabei bezeichnet $p(0)(w|h)$ einen geeigneten Initialisierungs- oder Anfangswert für die Wahrscheinlichkeit, dass ein Wort w auf eine bisherige Wortfolge h (Historie) folgt (S1/2).

20 In Verfahrensschritt S1/3 sind die aktuellen Iterationsrandwerte $m_\alpha^{(n)}$ für ihre jeweiligen zugehörigen gewünschten Randwerte m_α , welche letzten Endes das gewünschte Sprachmodell definieren, zu berechnen, und zwar für alle in dem Sprachmodell vorbestimmten Merkmale α .

25 Die gewünschten Randwerte m_α legen folgende Randbedingungen für die gesuchte Wahrscheinlichkeitsverteilung $p(w|h)$ fest:

$$\sum_{(h,w)} N(h) \cdot p(w|h) \cdot f_\alpha(h,w) = m_\alpha \quad (2)$$

wobei

- $N(h)$: eine Häufigkeit der Historie h ;
 $p(w|h)$: die Wahrscheinlichkeit, mit welcher das Wort w auf die
 5 Historie h folgt; und
 $f_\alpha(h, w)$: eine Merkmalsfunktion für das Merkmal α

bezeichnet.

- 10 Für das Schätzen geeigneter Randwerte sind im Stand der Technik verschiedene Ansätze bekannt.

Gemäß einem bekannten Ansatz ergibt sich der gewünschte Randwert m_α für das Sprachmodell durch Anwendung der Merkmalsfunktion f_α auf ein Trainingskorpus und

- 15 anschließendes Glätten der daraus resultierenden Häufigkeiten. Dabei kann das Glätten z.B. durch Subtraktion eines Korrekturwertes von der ermittelten Häufigkeit $N(\alpha)$ erfolgen.

- 20 Gemäß einer zweiten alternativen Methode erfolgt die Berechnung durch Reduktion von Merkmalsmengen in dem Sprachmodell so lange, bis die Randbedingungen keine Widersprüche mehr aufweisen. Eine derartige Reduktion von Merkmalsmengen muss in der Praxis sehr umfangreich sein, weil ansonsten das erzeugte Sprachmodell keine Lösung mehr zu der ursprünglichen Trainingsaufgabe darstellt.

- 25 Für die Merkmalsfunktion f_α sind ebenfalls verschiedene Definitionen im Stand der Technik bekannt; sie wird normalerweise jedoch definiert zu:

$$f_\alpha(h, w) = \begin{cases} 1 & \text{wenn } \alpha \text{ die Wortfolge } (h, w) \text{ korrekt beschreibt} \\ 0 & \text{sonst} \end{cases} \quad (3)$$

Der n'te Iterationsrandwert $m_\alpha^{(n)}$ stellt eine iterative Näherung für die soeben definierten gewünschten Randwerte m_α dar. Der n'te Iterationsrandwert $m_\alpha^{(n)}$ berechnet sich wie folgt:

$$m_\alpha^{(n)} = \sum_{(h,w)} N(h) \cdot p^{(n)}(w|h) \cdot f_\alpha(h,w) \quad (4)$$

Diese Formel unterscheidet sich von der oben genannten Formel (2) lediglich dadurch, dass für die Wahrscheinlichkeit $p(w|h)$ eine Approximation in Form des Iterationswertes $p^{(n)}(w|h)$ gewählt wird, wobei der Iterationswert $p^{(n)}$ wie folgt berechnet wird:

$$p^{(n)}(w|h) = \frac{1}{Z^{(n)}(h)} \cdot \exp\left(\sum_\alpha \lambda_\alpha^{(n)} \cdot f_\alpha(h,w)\right) \quad (5)$$

$$Z^{(n)}(h) = \sum_w \exp\left(\sum_\alpha \lambda_\alpha^{(n)} \cdot f_\alpha(h,w)\right) \quad (6)$$

- 15 wobei $Z^{(n)}(h)$ und die freien Parameter $\lambda_\alpha^{(n)}$ jeweils durch den GIS-Trainingsalgorithmus trainiert, d.h. iterativ approximiert werden.

- Gemäß Verfahrensschritt S1/4 ist nach jedem Iterationsschritt zu prüfen, ob die berechneten Iterationsrandwerte $m_\alpha^{(n)}$ bereits mit einer gewünschten Genauigkeit gegen die gewünschten Randwerte m_α konvergiert sind. Sollte dies der Fall sein, so ist das erfindungsgemäße Verfahren beendet.

- 20 Solange dies jedoch noch nicht der Fall ist, ist vor jeder Erhöhung des Iterationsparameters um 1 (wieder) die Merkmalsgruppe $A_i(n)$ mit dem größten Korrekturbedarf zu bestimmen; dies erfolgt erfindungsgemäß dadurch, dass die nachfolgend beschriebenen Verfahrensschritte S1/5 bis S1/7 erstmals bzw. wiederholt durchgeführt werden.

Gemäß Verfahrensschritt S1/5 wird für alle Merkmalsgruppen A_i in dem Sprachmodell jeweils getrennt das Kriterium $D_i^{(n)}$ berechnet, welches ein Maß dafür ist, inwieweit die Iterationswerte $m_\alpha^{(n)}$ für Merkmale α der Gruppe A_i an die jeweils zugehörigen gewünschten Randwerte m_α angepasst sind. Das Kriterium wird mathematisch vorzugsweise wie folgt beschrieben:

$$D_i^{(n)} = \left[\sum_{\alpha \in A_i} t_\alpha \cdot m_\alpha \log \left(\frac{m_\alpha}{m_\alpha^{(n)}} \right) + \left(1 - \sum_{\alpha \in A_i} t_\alpha \cdot m_\alpha \right) \log \left(\frac{1 - \sum_{\alpha \in A_i} t_\alpha \cdot m_\alpha}{1 - \sum_{\alpha \in A_i} t_\alpha \cdot m_\alpha^{(n)}} \right) \right] \quad (7)$$

Dabei berechnet sich die Konvergenzschriftweite t_α wie folgt:

10

$$t_\alpha = \frac{1}{M_i} \quad \text{mit} \quad M_i = \max_{(h,w)} \left(\sum_{\alpha \in A_i} f_\alpha(h,w) \right) \quad (8)$$

In bezug auf Formel (7) ist es wichtig zu erkennen, dass der Betrag von $D_i^{(n)}$ umso größer ist, je schlechter die Anpassung der iterativen Randwerte $m_\alpha^{(n)}$ an ihre zugehörigen gewünschten Randwerte m_α für eine bestimmte Merkmalsgruppe A_i ist.

15

Folglich ergibt sich die Merkmalsgruppe mit der schlechtesten Anpassung, d.h. mit dem größten Korrekturbedarf, gemäß Verfahrensschritt S1/7 zu:

$$i(n) = \arg \max_j D_j^{(n)} \quad (9)$$

20

In Verfahrensschritt S1/8 wird die so ausgewählte Merkmalsgruppe $A_{i(n)}$ mit dem größten Korrekturbedarf zur Berechnung der $n+1$ Iterationswerte für die freien Parameter λ z.B. gemäß der aus dem Stand der Technik bekannten und oben beschriebenen Gleichung (1) verwendet. Dabei wird während des n -ten Iterationsschrittes die Gleichung (1) für alle Merkmale α aus der ausgewählten Merkmalsgruppe $A_{i(n)}$ berechnet, bevor der Iterations-

25

parameter n um 1 erhöht wird. Die Iterationswerte $\lambda_\alpha^{(n+1)}$ berechnen sich dann gemäß einem ersten Ausführungsbeispiel einer mathematischen Funktion $G()$ zu:

$$\lambda_\alpha^{(n+1)} = G() = \lambda_\alpha^{(n)} + t_\alpha \cdot \log \left(\frac{m_\alpha}{m_\alpha^{(n)}} \cdot \frac{1 - \sum_{\beta \in Ai(n)} t_\beta \cdot m_\beta^{(n)}}{1 - \sum_{\beta \in Ai(n)} t_\beta \cdot m_\beta} \right) \quad (10)$$

5

Eine derartige azyklische Berechnung der Iterationswerte $\lambda_\alpha^{(n+1)}$ bietet den Vorteil, dass unnötige Iterationsschritte vermieden werden, und dass auf diese Weise die Konvergenzgeschwindigkeit des GIS-Trainingsalgorithmus erheblich verbessert wird.

- 10 Nach der Berechnung des Iterationswertes $\lambda_\alpha^{(n+1)}$ erfolgt in Schritt S1/8 eine Neudefinition des Iterationsparameters n zu $n = n+1$.

Der gemäß Formel (8) berechnete und gemäß Schritt S1/8 von $\lambda_\alpha^{(n+1)}$ nach $\lambda_\alpha^{(n)}$ umdefinierte Iterationswert dient dann wieder zur Berechnung des aktuellen Iterationsrandwertes

- 15 $m_\alpha^{(n)}$ in Schritt S1/3 gemäß Formel (4) in Verbindung mit den Formeln (5) und (6).

Die Konvergenzgeschwindigkeit des GIS-Trainingsalgorithmus hängt allerdings nicht nur von der Auswahl einer geeigneten Merkmalsgruppe für jeden Iterationsschritt n ab, sondern auch von der Merkmalsfunktion, mit welcher die Konvergenzschrittweiten t_α und

- 20 t_β sowie die iterativen Randwerte $m_\alpha^{(n)}$ und $m_\beta^{(n)}$ berechnet werden. So kann die Konvergenzgeschwindigkeit des GIS-Trainingsalgorithmus zusätzlich dadurch gesteigert werden, dass anstelle einer normalen Merkmalsfunktion gemäß Formel (3) eine orthogonalisierte

Merkmalsfunktion f_α^{ortho} verwendet wird, welche folgendermaßen definiert ist:

25

$$f_\alpha^{ortho}(h, w) = \begin{cases} 1 & \text{wenn } \alpha \text{ das Merkmal mit der höchsten Reichweite} \\ & \text{in } Ai \text{ ist, welches die Wortfolge } (h, w) \text{ korrekt} \\ & \text{beschreibt} \\ 0 & \text{sonst} \end{cases} \quad (11)$$

- Bei Einsetzen der orthogonalisierten Merkmalsfunktion f_{α}^{ortho} anstelle der normalen Merkmalsfunktion f_{α} in die Formeln (4), (5), (6) und (8) und bei zusätzlicher Berechnung der gewünschten Randwerte m_{α}^{ortho} und m_{β}^{ortho} durch Anwenden der orthogonalisierten Merkmalsfunktion f_{α}^{ortho} auf ein Trainingskorpus des Sprachmodells ergibt sich die
- 5 Formel für den GIS-Trainingsalgorithmus in Analogie zu Formel (10) zu:

$$\lambda_{\alpha}^{ortho(n+1)} = \lambda_{\alpha}^{ortho(n)} + t_{\alpha}^{ortho} \cdot \log \left(\frac{m_{\alpha}^{ortho}}{m_{\alpha}^{ortho(n)}} \cdot \frac{1 - \sum_{\beta \in A_i(n)} t_{\beta}^{ortho} \cdot m_{\beta}^{ortho(n)}}{1 - \sum_{\beta \in A_i(n)} t_{\beta}^{ortho} \cdot m_{\beta}^{ortho}} \right), \quad (12)$$

- wobei auch die Berechnung dieser Formel vorzugsweise nicht zyklisch, sondern azyklisch gemäß dem in Fig. 1 beschriebenen Verfahren erfolgt. Die rechte Seite der Gleichung (12) beschreibt ein zweites Ausführungsbeispiel für die mathematische Funktion G in Patent-
- 10 anspruch 1.

- Die Berechnung der gewünschten Randwerte m_{α}^{ortho} , welche durch die Iterationswerte $m_{\alpha}^{ortho(n)}$ approximiert werden sollen, erfolgt vorzugsweise gemäß:
- 15

$$m_{\alpha}^{ortho} = m_{\alpha} - \sum_{(*)} m_{\beta}^{ortho} \quad (13)$$

- wobei (*) alle höher-reichweitigen Merkmale β umfasst, welche das Merkmal α einschließen und welche in derselben Merkmalsgruppe liegen wie α . Zur Berechnung des
- 20 Randwertes m_{β}^{ortho} ist die genannte Formel quasi rekursiv für jedes Merkmal β immer wieder anzuwenden, bis für bestimmte Merkmale, nämlich für jene mit der höchsten Reichweite, der Summenterm verschwindet, weil zu diesen Merkmalen keine höherreichweitigen Merkmale existieren. Die gewünschten orthogonalisierten Randwerte für die
- 25 höchstreichweitigen Merkmale β_k entsprechend dann jeweils den normalen gewünschten Randwerten $m_{\beta k}$.

Ein Verfahren zur Berechnung der gewünschten orthogonalisierten Randwerte m_{α}^{ortho} gemäß Formel (13) ist in Fig. 2a und 2b beschrieben.

5 Gemäß den Fig. 2a und 2b werden in einem ersten Verfahrensschritt S2/1 in dem Sprachmodell alle Merkmale β_i mit $i = 1 \dots g$ bestimmt, welche eine sog. höhere Reichweite aufweisen als ein Merkmal $\alpha = \beta_0$, d.h. welche dieses an einer vorbestimmten Stelle einschließen und welche aus derselben Merkmalsgruppe stammen wie α . Anschließend wird in einem Verfahrensschritt S2/2 für alle Merkmale β_i mit $i = 0 \dots g$, also auch für das Merkmal $\alpha = \beta_0$, ein gewünschter Randwert m_{β_i} berechnet.

10

Für die Berechnung des gewünschten Randwertes m_{β_i} sind verschiedene Methoden im Stand der Technik bekannt, wie oben nachfolgend zu Formel (2) beschrieben wurde.

In Verfahrensschritt S2/3 werden nachfolgend alle Merkmale β_i nach ihrer Reichweite
15 sortiert, wobei vorzugsweise dem Merkmal β_i mit der größten Reichweite der Index $i = g$ zugeordnet wird. Dabei kann es durchaus vorkommen, dass einzelnen Reichweiteklassen, also z.B. der Klasse Bigramme oder der Klasse Trigramme, mehrere Merkmale β_i zugeordnet werden. In diesen Fällen sind mehrere Merkmale β_i mit unterschiedlichen, aber aufeinanderfolgenden Indizes i ein und derselben Reichweiteklasse zugeordnet, d.h. diese
20 Merkmale haben dann jeweils dieselbe Reichweite.

Für den Ablauf des Verfahrens, bei welchem in den nachfolgenden Schritten die einzelnen Merkmale β_i der Reihe nach ausgewertet werden, ist es wichtig, dass in einem ersten Durchlauf $n=0$ des Verfahrens mit einem Merkmal β_i begonnen wird, welches der
25 höchsten Reichweitenklasse zugeordnet ist; vorzugsweise wird deshalb mit dem Merkmal β_g begonnen (siehe Verfahrensschritte S2/4 und S2/5 in Fig. 2a).

In einem nachfolgenden Verfahrensschritt S2/6 wird geprüft, ob es zu dem aktuell ausgewählten Merkmal β_i (beim ersten Durchlauf $n = 0$ ist $i = g$), vorbestimmte höherreichweitigere Merkmale β_k mit $i < k < g$ gibt, welche das Merkmal β_i einschließen. Beim
30

- ersten Durchlauf gehört das Merkmal β_i , wie oben gesagt, automatisch der Klasse mit der höchsten Reichweite an und deshalb ist die Abfrage in Verfahrensschritt S2/6 für dieses Merkmal β_i zu verneinen. In diesem Fall springt das Verfahren zu Verfahrensschritt S2/8, wo ein Parameter X zu Null gesetzt wird. Es erfolgt daraufhin eine Berechnung eines verbesserten gewünschten orthogonalisierten Randwertes m_{β}^{ortho} für das Merkmal β_i gemäß Verfahrensschritt S2/9 in Fig. 2b. Wie dort ersichtlich, wird dieser Randwert für das Merkmal β_i dem in Schritt S2/2 berechneten gewünschten Randwert m_{β_i} gleichgesetzt, wenn der Parameter X=0 ist.
- 5
- 10 Die Verfahrensschritte S2/5 bis S2/11 werden daraufhin sukzessive für alle Merkmale β_{i-1} mit $i-1=g-1, g-2, \dots, 0$ wiederholt. In Verfahrensschritt S2/10 findet eine dafür notwendige Neuinitialisierung des Indexes i statt und in Verfahrensschritt S2/11 erfolgt eine Abfrage, ob alle Merkmale β_i mit $i = 0 \dots g$ abgearbeitet worden sind.
- 15 Für alle Merkmale β_i , für welche vorbestimmte höher-reichweitige Merkmale β_k mit $i < k \leq g$ existieren, ist die Abfrage in Verfahrensschritt S2/6 mit "Ja" zu beantworten. Der Parameter X wird dann nicht zu Null gesetzt, sondern berechnet sich gemäß Verfahrensschritt S2/7 durch Aufsummierung der entsprechenden, in vorherigen Durchläufen jeweils in Verfahrensschritt S2/9 berechneten verbesserten gewünschten orthogonalisierten
- 20 Randwerte $m_{\beta_k}^{ortho}$ für die jeweils höher-reichweitigen Merkmale β_k .
- Sobald in Verfahrensschritt S2/11 festgestellt worden ist, dass der gewünschte orthogonalisierte Randwert $m_{\beta_0}^{ortho}$ in Verfahrensschritt S2/9 berechnet worden ist, wird dieser in Verfahrensschritt S2/12 als m_{α}^{ortho} ausgegeben.
- 25 In Fig. 3 ist schließlich ein Spracherkennungssystem 10 gemäß der vorliegenden Erfindung dargestellt, welches auf Basis des sog. Maximum-Entropie-Sprachmodells arbeitet. Es umfasst eine Erkennungseinrichtung 12, welche den Bedeutungsinhalt von bereitgestellten Sprachsignalen zu erkennen versucht. Die Sprachsignale werden dem Spracherkennungs-

system üblicherweise als Ausgangssignale eines Mikrophons 20 bereitgestellt. Die Erkennungseinrichtung 12 erkennt den Bedeutungsinhalt der Sprachsignale dadurch, dass sie Muster in dem empfangenen akustischen Signal auf vordefinierte Erkennungssymbole, wie z.B. bestimmte Wort, Handlungen oder Ereignisse, mit Hilfe des implementierten Maximum Entropie Sprach Modells MESM abbildet. Schließlich gibt die Erkennungseinrichtung 12 ein Ausgangssignal aus, welches den in dem Sprachsignal erkannten Bedeutungsinhalt repräsentiert und zur Ansteuerung von verschiedensten Geräten, z.B. eines Textverarbeitungsprogramms oder eines Telefons nach Maßgabe durch diesen Bedeutungsinhalt dienen kann.

10

Für eine möglichst fehlerfreie Ansteuerung der Geräte im Sinne des Bedeutungsinhaltes von steuernder Sprachinformation ist es erforderlich, dass das Spracherkennungssystem 10 die Bedeutungsinhalte der auszuwertende Sprache mit einer möglichst hohen Quote richtig erkennt. Dazu ist eine möglichst gute Anpassung Sprachmodells an die sprachlichen Besonderheiten des Sprechers, d.h. des Benutzers des Spracherkennungssystems erforderlich.

15

Diese Anpassungsaufgabe leistet eine Trainingseinrichtung 14, welche entweder extern oder in das Spracherkennungssystem 10 integriert betrieben werden kann. Genauer gesagt dient die Trainingseinrichtung 14 zum Anpassen des MESM in dem Spracherkennungssystem 10 an wiederkehrende statistische Muster in der Sprache eines bestimmten

20

Benutzers.

Sowohl die Erkennungseinrichtung 12 wie auch die Trainingseinrichtung 14 sind üblicherweise, aber nicht notwendigerweise, als Softwaremodule ausgebildet und laufen auf einem geeigneten Computer (nicht gezeigt) ab.

25

PATENTANSPRÜCHE

1. Verfahren zum Berechnen von Iterationswerten für freie Parameter $\lambda\alpha$ in dem Maximum Entropie Sprachmodell gemäß dem folgenden allgemeinen Trainingsalgorithmus:

$$5 \quad \lambda_{\alpha}^{(n+1)} \Big|_{\alpha \in A_i(n)} = G(\lambda_{\alpha}^{(n)}, m_{\alpha}, m_{\alpha}^{(n)}, \dots) \Big|_{\alpha \in A_i(n)}$$

wobei:

- n : einen Iterationsparameter, welcher einen aktuellen Iterationsschritt repräsentiert;
- 10 A_i : die i'te Merkmalsgruppe in dem Sprachmodell mit $1 \leq i \leq m$;
- $A_i(n)$: die in dem n'ten Iterationsschritt ausgewählte Merkmalsgruppe;
- α : ein Merkmal in dem Sprachmodell;
- G : eine mathematische Funktion;
- 15 $\lambda_{\alpha}^{(n)}$: den n'ten Iterationswert für den freien Parameter $\lambda\alpha$;
- m_{α} : einen gewünschter Randwert für das Merkmal α ; und
- $m_{\alpha}^{(n)}$: den n'ten Iterationsrandwert für den gewünschten Randwert $m\alpha$;
- 20 bezeichnet,

wobei jedem Iterationsparameter n eine Merkmalsgruppe $A_i(n)$ von insgesamt m Merkmalsgruppen des Sprachmodells zugeordnet ist, und wobei die Iterationswerte $\lambda_\alpha^{(n+1)}$ jeweils für alle Merkmale α aus der aktuell zugeordneten Merkmalsgruppe $A_i(n)$ berechnet werden;

5 dadurch gekennzeichnet,

dass dem aktuellen Iterationsparameter n jeweils diejenige Merkmalsgruppe $A_i(n)$ mit $1 \leq i(n) \leq m$ zugeordnet wird, für welche, gemäß einem vorbestimmten Kriterium, die Anpassung der Iterationsrandwerte $m_\alpha^{(n)}$ an die jeweils zugehörigen gewünschten Randwerte m_α am schlechtesten im Vergleich zu allen m Merkmalsgruppen des

10 Sprachmodells erfüllt ist.

2. Verfahren nach Anspruch 1,

dadurch gekennzeichnet,

dass es vor jeder Erhöhung des Iterationsparameters n folgende Schritte zur Berechnung und Auswertung des Kriteriums umfasst:

a) Berechnen von aktuellen Iterationsrandwerten $m_\alpha^{(n)}$ für die Merkmale α aller Merkmalsgruppen A_i mit $1 \leq i \leq m$ des Sprachmodells gemäß der folgenden Formel:

$$m_\alpha^{(n)} = \sum_{(h,w)} N(h) \cdot p^{(n)}(w|h) \cdot f_\alpha(h,w)$$

20

wobei $N(h)$: die Häufigkeit beschreibt, mit welcher die Wortfolge h (Historie) in einem Trainingskorpus des Sprachmodells vorkommt;

25 $p^{(n)}(w|h)$: ein Iterationswert für die Wahrscheinlichkeit ist, mit welcher das Wort w auf die Historie h folgt; und

$f_\alpha(h,w)$: eine Merkmalsfunktion für das Merkmal α repräsentiert;

b) Auswählen derjenigen Merkmalsgruppe $A_i(n)$, für welche die Iterationsrandwerte $m_\alpha^{(n)}$ am schlechtesten an die zugehörigen Randwerte m_α angepasst sind durch Ausführen der folgenden Schritte:

- 5 bi) für jede Merkmalsgruppe A_i : Berechnen des Kriteriums

$D_i^{(n)}$ gemäß folgender Formel:

$$D_i^{(n)} = \left[\sum_{\alpha \in A_i} t_\alpha \cdot m_\alpha \cdot \log \left(\frac{m_\alpha}{m_\alpha^{(n)}} \right) + \left(1 - \sum_{\alpha \in A_i} t_\alpha \cdot m_\alpha \right) \cdot \log \left(\frac{1 - \sum_{\alpha \in A_i} t_\alpha \cdot m_\alpha}{1 - \sum_{\alpha \in A_i} t_\alpha \cdot m_\alpha^{(n)}} \right) \right];$$

- 10 bii) Auswählen der Merkmalsgruppe $A_i(n)$ mit dem größten Wert für das Kriterium $D_i^{(n)}$ gemäß:

$$i(n) \equiv \arg \max_j D_j^{(n)} \quad (7);$$

biii) Aktualisieren des Parameters $\lambda_\alpha^{(n+1)}$ für alle Merkmale α aus der ausgewählten Merkmalsgruppe $A_i(n)$; und

- 15 c) Wiederholen der Schritte a) und b) in jedem weiteren Iterationsschritt, solange bis alle Randwerte $m_\alpha^{(n+1)}$ eine gewünschte Konvergenzgenauigkeit erreicht haben.

3. Verfahren nach Anspruch 2,

dadurch gekennzeichnet,

- 20 dass vor der erstmaligen Durchführung der Schritte a) - c) gemäß Anspruch 2 folgende Initialisierungsschritte durchgeführt werden:

a') Bestimmen von Werten für die Konvergenzschrittweiten α ; und

a'') Initialisieren von $p(0)(w|h)$ mit einem beliebigen Satz von Parametern $\lambda_\alpha^{(0)}$.

4. Verfahren nach Anspruch 3,

dadurch gekennzeichnet,

- 5 dass in Schritt a') die Werte der Konvergenzschriftweiten t_α für jede Merkmalsgruppe A_i wie folgt berechnet werden:

$$t_\alpha = \frac{1}{M_i} \text{ mit } M_i = \max_{(h,w)} \left(\sum_{\alpha \in A_i} f_\alpha(h,w) \right)$$

- 10 5. Verfahren nach einem der vorangegangenen Ansprüche,

dadurch gekennzeichnet,

dass die Funktion G einen Generalized Iterativ Scaling GIS-Trainingsalgorithmus repräsentiert und folgendermaßen definiert ist:

15
$$\lambda_\alpha^{(n+1)} = G = \lambda_\alpha^{(n)} + t_\alpha \cdot \log \left(\frac{m_\alpha}{m_\alpha^{(n)}} \cdot \frac{1 - \sum_{\beta \in A_i(n)} t_\beta \cdot m_\beta^{(n)}}{1 - \sum_{\beta \in A_i(n)} t_\beta \cdot m_\beta} \right),$$

wobei α ein bestimmtes und β alle Merkmale aus der ausgewählten Merkmalsgruppe $A_i(n)$ bezeichnet.

- 20 6. Verfahren nach einem der Ansprüche 2 bis 5,

dadurch gekennzeichnet,

dass die Merkmalsfunktion f_α eine orthogonalisierte Merkmalsfunktion f_α^{ortho} ist, welche wie folgt definiert ist:

25
$$f_\alpha^{ortho}(h,w) = \begin{cases} 1 & \text{wenn } \alpha \text{ das Merkmal mit der höchsten Reichweite} \\ & \text{in } A_i \text{ ist, welches die Wortfolge } (h,w) \text{ korrekt} \\ & \text{beschreibt} \\ 0 & \text{sonst} \end{cases}$$

7. Verfahren nach Anspruch 6,

dadurch gekennzeichnet,

dass der gewünschte orthogonalisierte Randwert m_{α}^{ortho} berechnet wird gemäß:

$$m_{\alpha}^{ortho} = m_{\alpha} - \sum_{(*)} m_{\beta}^{ortho}$$

5

wobei (*) alle höher-reichweitigen Merkmale β umfasst, welche das Merkmal α einschließen und welche aus derselben Merkmalsgruppe stammen wie α .

10 8. Spracherkennungssystem (10) mit:

einer Erkennungseinrichtung (12) zum Erkennen des Bedeutungsinhaltes eines von einem Mikrophon (20) aufgenommenen und bereitgestellten akustischen Signals, insbesondere eines Sprachsignals, durch Abbilden von Teilen dieses Signals auf vordefinierte Erkennungssymbole, wie sie von dem implementierten Maximum Entropie Sprach Modell

15 MESM angeboten werden, und zum Erzeugen von Ausgangssignalen, welche den erkannten Bedeutungsinhalt repräsentieren; und

einer Trainingseinrichtung (14) zum Anpassen des MESM an wiederkehrende statistische Muster in der Sprache eines bestimmten Benutzers des Spracherkennungssystems (10);

dadurch gekennzeichnet,

20 dass die Trainingseinrichtung (14) freie Parameter λ in dem MESM gemäß dem Verfahren nach Anspruch 1 berechnet.

9. Trainingseinrichtung (14) zum Anpassen des Maximum Entropie Sprach Modells

MESM in einem Spracherkennungssystem (10) an wiederkehrende statistische Muster in

25 der Sprache eines bestimmten Benutzers des Spracherkennungssystems (10),

dadurch gekennzeichnet,

dass die Trainingseinrichtung (14) freie Parameter λ in dem MESM gemäß dem Verfahren nach Anspruch 1 berechnet.

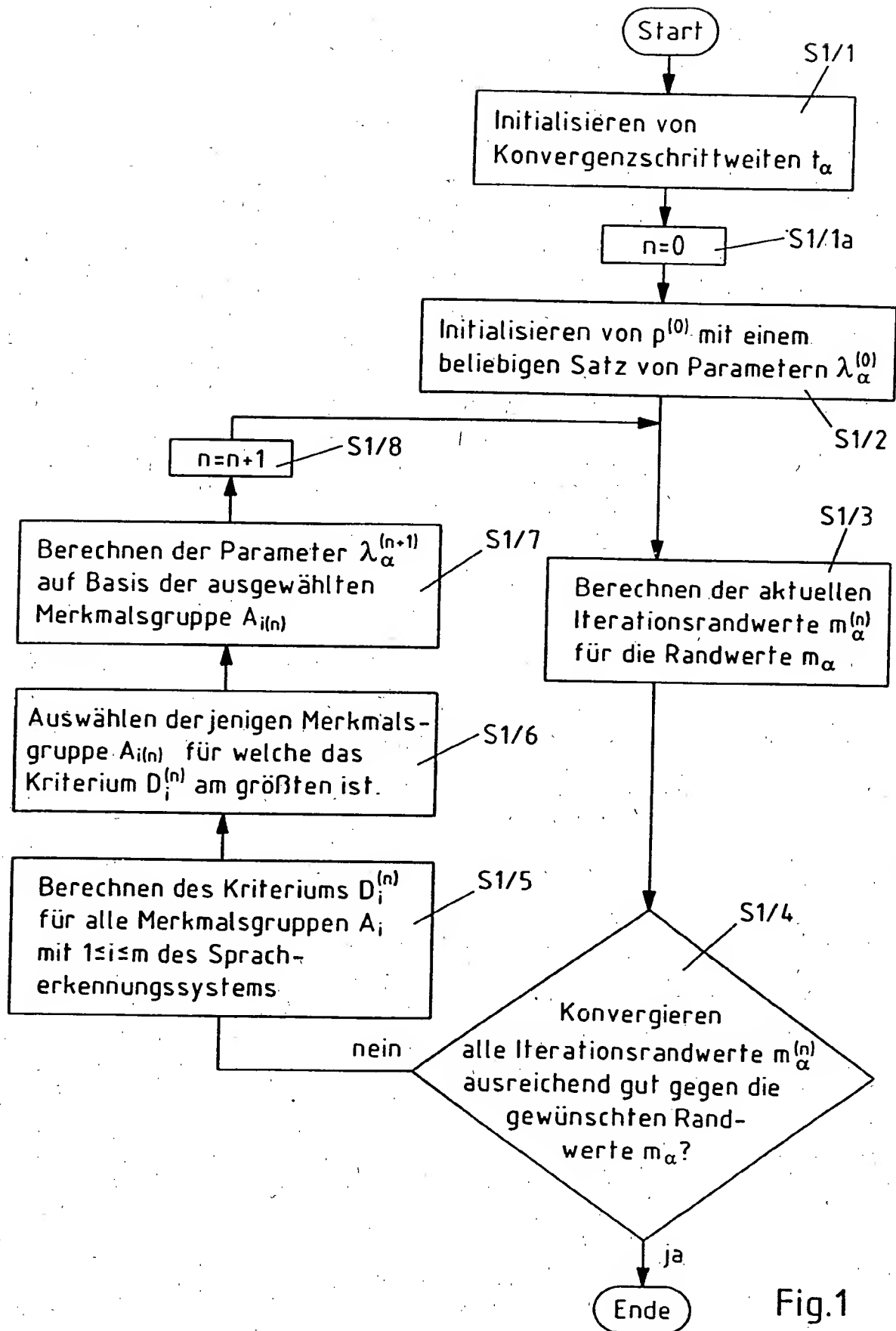


Fig.1

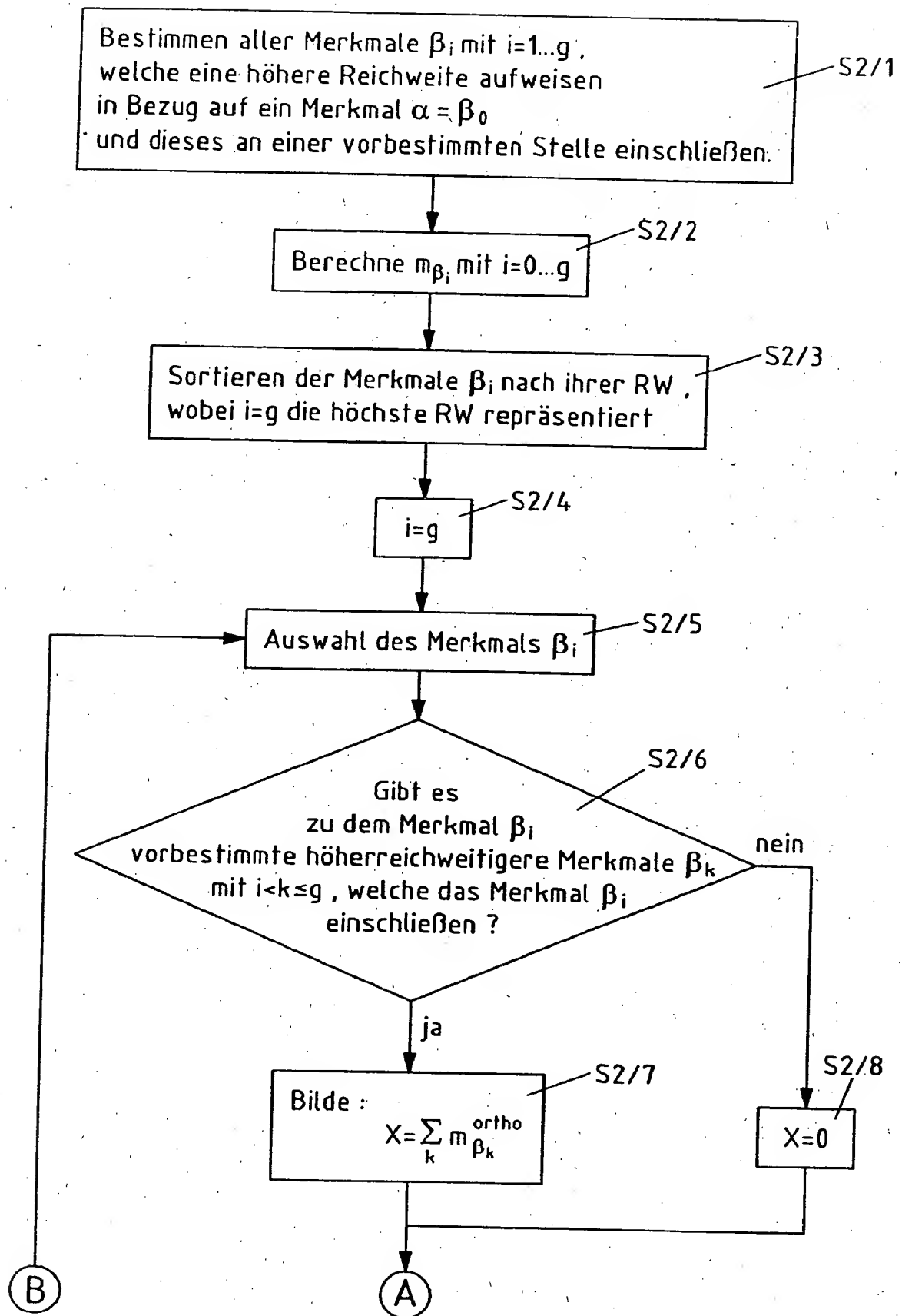


Fig.2a

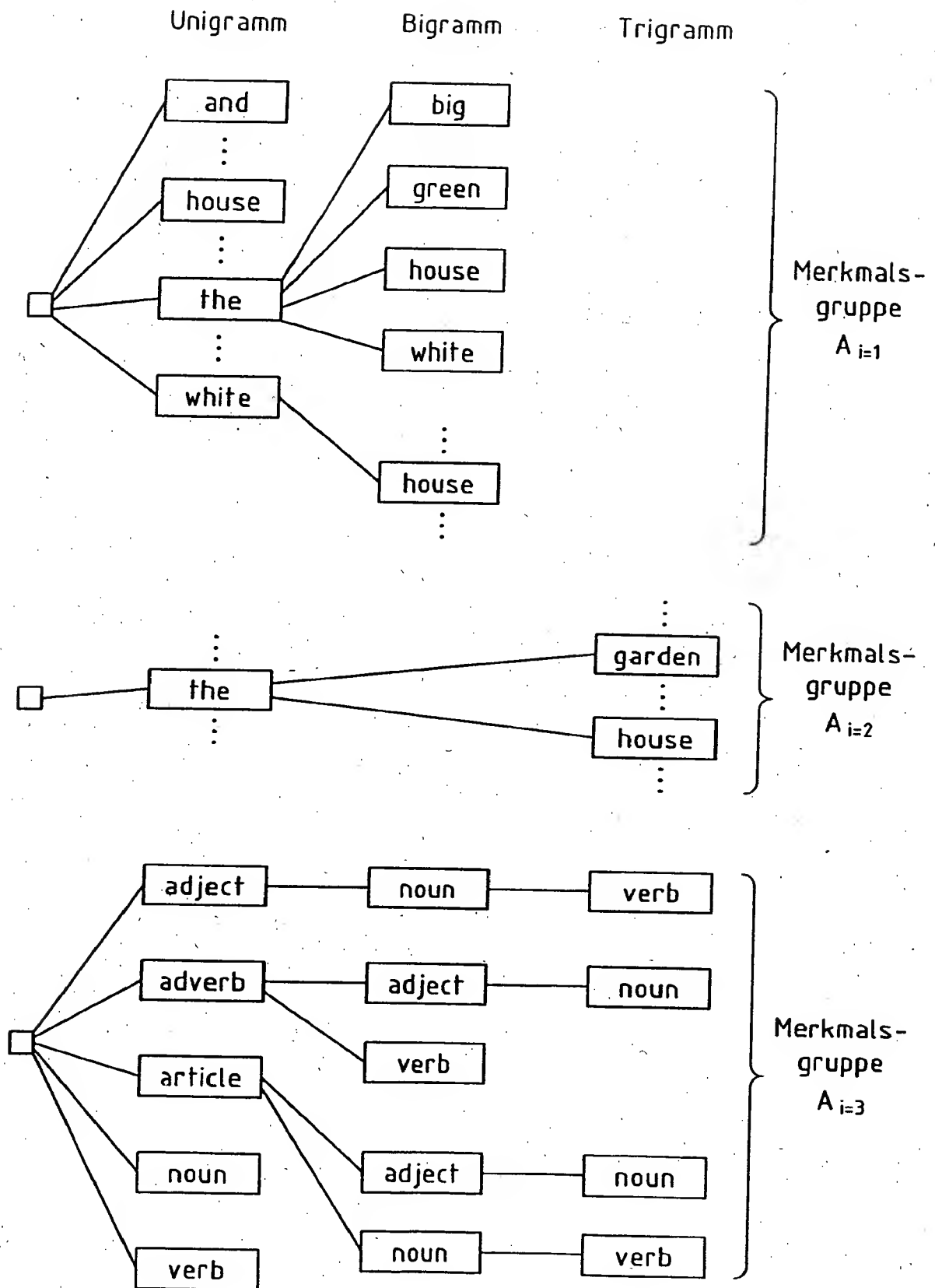
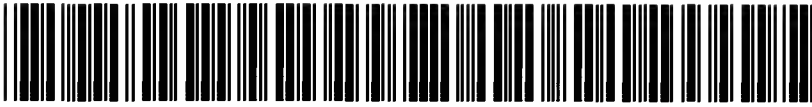


Fig.4 Stand der Technik



Creation date: 12-03-2004

Indexing Officer: HABDULEKADER - HUDA ABDULEKADER

Team: OIPEScanning

Dossier: 10187363

Legal Date: 08-06-2002

No.	Doccode	Number of pages
1	FRPR	13

Total number of pages: 13

Remarks:

Order of re-scan issued on